Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 14, Issue Number 1 (1996)

Design of Sequence-Specific DNA Binding Ligands That Use a Two-Stranded Peptide Motif for DNA Sequence Recognition

http://www.albany.edu/chemistry/sarma/jbsd.html
http://www.albany.edu/chemistry/sarma/shafer.html

V. A. Nikolaev[1], S. L. Grokhovsky[1], A. N. Surovaya[1], T. A. Leinsoo[1], N. Yu. Sidorova[1], A. S. Zasedatelev[1], A. L. Zhuze[1], G. A. Strahan[2], R. H. Shafer[2]*, and G. V. Gursky[1]*

[1]Engelhardt Institute of
Molecular Biology,
Russian Academy of Sciences,
Moscow 117984, Russia

[2]Department of
Pharmaceutical Chemistry,
University of California,
San Francisco, CA 94143, USA

Abstract

The design and DNA binding activity of β-structure-forming peptides and netropsin-peptide conjugates are reported. It is found that a pair of peptides - S,S'-bis(Lys-Gly-Val-Cys-Val-NH-NH-Dns) - bridged by an S-S bond binds at least 10 times more strongly to poly(dG)•poly(dC) than to poly(dA)•poly(dT). This peptide can also discriminate between 5'-GpG-3' and 5'-GpC-3' steps in the DNA minor groove. Based on these observations, new synthetic ligands, bis-netropsins, were constructed in which two netropsin-like fragments were attached by means of short linkers to a pair of peptides - Gly-Cys-Gly- or Val-Cys-Val - bridged by S-S bonds. These compounds possess a composite binding specificity: the peptide chains recognize 5'-GpG-3' steps on DNA, whereas the netropsin-like fragments bind preferentially to runs of 4 AT base pairs. Our data indicate that combining the AT-base-pair specific properties of the netropsin-type structure with the 5'-GpG-3'-specific properties of certain oligopeptides offers a new approach to the synthesis of ligands capable of recognizing mixed sequences of AT- and GC-base pairs in the DNA minor groove. These compounds are potential models for DNA-binding domains in proteins which specifically recognize base pair sequences in the minor groove of DNA.

Introduction

There is much interest in the design and synthesis of compounds that possess binding specificity comparable to that of bacterial repressors. These compounds could be used to selectively regulate the activity of certain genes in bacterial and eukaryotic cells and might have important applications in pharmacology. One approach to this problem is based on the use of antibiotics of the distamycin class. It is well known that distamycin A and netropsin (Nt) bind in the minor groove of DNA and exhibit a pronounced specificity for AT-rich regions (1-7). It is of interest to construct DNA-binding ligands that will exhibit greater binding specificity than distamycin A and netropsin, recognizing a mixed sequence of AT and GC base pairs.

Based on the X-ray structure of the 1:1 complex of netropsin with a self-complementary DNA oligomer (8-10), attempts were undertaken to synthesize ligands capable of recognizing GC-containing sequences through specific hydrogen bonding to the guanine amino group in the minor groove (11-13). It was shown that distamycin analogs in which the pyrrole CH was replaced with a heteroatom capable of forming a hydrogen bond to the guanine 2-amino group displayed increased tolerance for GC base pairs in their binding sites, with an overall loss in specificity (11-13). Recent studies on DNA-binding mechanisms of distamycin analogs have revealed a new structural motif consisting of two antibiotic molecules packed in the minor groove in an antiparallel, side-by-side manner. This binding mechanism was first proposed by Burkhardt et. al. (14), but structural details became available from

*Authors to whom correspondence should be addressed. G. V. Gursky- Phone: (7)-095-135- 9790; Fax: (7)-095-938-2187; E-mail: gursky@imb.imb.ac.ru. R. H. Shafer- Phone: (415)-476-2761; Fax: (415)-476-0688; E-mail: shafer@cgl.ucsf.edu.

NMR studies of Wemmer and co-workers on the binding of distamycin A to a DNA oligomer (15,16). Very recently, the crystalline structure of a 2:1 complex between distamycin A and an inosine-containing alternating duplex, $d(ICICICIC)_2$, has been determined (17). In this complex, the two distamycin molecules are sandwiched in the minor groove, with the pyrrole rings of one antibiotic molecule stacking on the peptide groups of the other. Each bound antibiotic molecule forms hydrogen bonds with cytosine O2 and inosine N3 atoms lying in the neighboring oligonucleotide strand.

The side-by-side dimer motif was used by Dervan and coworkers in the elegant design of new sequence-specific DNA-binding ligands containing pyridine-2-carboxamide or 1-methylimidazole-2-caboxamide linked to the N terminus of bis-N-methylpyrrole-carboxamide (18,19). As revealed by footprinting analysis and NMR studies, these compounds recognize the sequence 5'-TGACT-3' as side-by-side dimers (19). Simultaneously, Lown and coworkers have characterized by NMR a 2:1 complex formed by a distamycin analog containing an imidazole ring at the central part of the molecule (20,21). Very recently Geierstanger et al. (22) have reported that a four-ring analog, with the sequence imidazole-pyrrole-imidazole-pyrrole, binds preferentially to DNA regions with a sequence 5'-(A,T)GCGC(A,T)-3' as a dimer.

Dimeric motifs are not limited by N-methylpyrrolecarboxamide analogs. In our previous work (26-30) we have shown that certain small peptides with inert amino acid side chains (such as trivaline) bind, in a β-associated form, more tightly to poly(dG)•poly(dC) than to poly(dA)•poly(dT). Attachment of trivaline to the amino terminus of a netropsin analogue alters its binding specificity (30). It has been shown that netropsin-peptide conjugates containing trivaline self-associate in aqueous solution and form dimers, which are in concentration-dependent equilibrium with the monomers. The binding specificity shown by a netropsin-peptide conjugate depended on the aggregation state of the oligopeptide. The monomer, like netropsin itself, binds more strongly to poly(dA)•poly(dT) than to any naturally occurring DNA, including calf thymus DNA or *Clostridium perfringens* DNA. In contrast, the netropsin-peptide conjugates, in a self-associated form, bind more strongly to calf thymus and *Clostridium perfringens* DNA than to poly(dA)•poly(dT) (30). Footprinting analysis has revealed that the highest affinity binding sites for dimeric species of netropsin-trivaline conjugates contain two runs of four AT base pairs separated by a GG-step (31).

In the present work, we describe the DNA binding properties of two sequence-specific ligands composed of a pair of peptides - Lys-Gly-Val-Cys-Val-NH-NH-Dns - (peptide **I**) or -Gly-Cys-Gly-NH-NH-Dns- (peptide **II**) bridged by a disulfide bond (Figure 1). Here Dns is a residue of 5-diaminonaphthalene-1-sulfonic acid. We have found, for example, that peptide **I** binds more strongly to poly(dG)•poly(dC) than to poly(dA)•poly(dT). It can also discriminate between 5'-GpG-3' and 5'-GpC-3' sequences on DNA. Our next step was to construct netropsin-peptide conjugates containing a double-stranded peptide motif. We have synthesized bis-netropsins in which two netropsin-like fragments are linked covalently to a pair of peptides, -Gly-Cys-Gly- or -Val-Cys-Val- , bridged by disulfide bonds (31). We present data showing that these compounds exhibit a composite binding specificity: the peptide chains recognize 5'-GpG-3' steps on DNA, whereas the netropsin-like fragments bind preferentially to runs of 4 AT base pairs.

Materials and Methods

Synthetic ligands and polynucleotides. Synthesis of peptides **I** and **II** (see Figure 1), the netropsin analog containing two glycine residues at the N-terminus and the bis-netropsins (Figures 6 and 7) was carried out as described in our previous publications (29-31). Lyophilized peptides **I** and **II** and bis-netropsins were dissolved in 1 mM Na-cacodylate (pH 7) immediately before use; their concentrations were

determined spectrophotometrically using the molar extinction coefficients ε_{330} = 4300, ε_{297} = 20000, and ε_{297} = 42000 $M^{-1}cm^{-1}$ for solutions containing peptide, netropsin analog and bis-netropsins, respectively.

For our binding studies, we used calf thymus DNA (ε_{259}= 13,300) and *Escherichia coli* DNA (ε_{259} = 13, 500) from Sigma, *Clostridium perfringence* DNA (ε_{259} = 12,000) and *Micrococus lysodeikticus* DNA (ε_{259} =14,000) from Pharmacia, poly(dA)•poly (dT) (ε_{259} = 12,000), poly(dG)•poly(dC) (ε_{259} = 14,800) and poly[d(GC)]•poly[d(GC)] (ε_{259} = 16,800) from PL Biochemicals (all extinction coefficients are given per mole of base pairs). Poly(dG)•poly(dC) was dissolved in 0.1 N NaOH and dialyzed for 24 h against the same solution with 1 mM EDTA. Before use, all polynucleotides were dialyzed against two changes of 1 mM Na-cacodylate buffer pH 7.0 without EDTA. Their CD spectra were in agreement with those in the literature. Poly(dG)•poly(dC) exhibited slow changes in the shape of its CD spectra with storage at 4 °C; in cases where CD spectra were markedly changed, the pH was adjusted to a value greater than 11 and dialysis was repeated.

Spectroscopic methods. Steady-state fluorescence measurements were performed on an Aminco SPF-1000 Sc instrument with excitation at 380 nm and emission at 520 nm (slits were 10 nm and 20 nm for excitation and emission, respectively). UV spectra were recorded on a Specord M40 and CD spectra on a Jobin Yvon Mark III dichrograph using 0.1, 0.5 , 1.0 and 5 cm pathlength cells. Binding constants of the netropsin analog and bis-netropsins to various naturally occurring and synthetic DNA's were determined as reported earlier using circular dichroism (CD) spectroscopy (Leinsoo et al., 1989). Binding was monitored at 315 nm, where free nucleic acid exhibits no CD signal. The molar dichroism of complexes between ligand and various DNA's was determined in a separate series of experiments performed in the presence of a large molar excess of nucleic acid over synthetic ligand. The molar dichroism of the free netropsin analog is equal to zero, whereas bis-netropsins free in solution were found to be optically active with a molar dichroism at 315 nm varying from 0.6 to 4.0 for various synthetic ligands (31).

DNA footprinting. Isolation of DNA fragment A (214 base pair long) involved EcoRI and HpaII (Boehringer Mannheim) cleavage of a modified pUC9 plasmid carrying a phage 434 G fragment cloned at the EcoRI site (32). The upper and lower 3' ends were labeled with [d-^{32}P]dATP and [d-^{32}P]dCTP (Izotop, Tashkent), respectively, and the Klenow fragment (Boehringer Mannheim), as described by Sambrook et al. (33). DNA fragment B (223 base pairs) is part of a human urokinase gene (34), isolated in 5% polyacrylamide gel (PAG) (32) after cleavage with TaqI and DdeI (Pharmacia) of a plasmid containing the cloned human urokinase gene. The plasmid was kindly donated by Dr. R. Sh. Beabealaschvilli.

DNase footprinting was performed according to standard protocols (35). Details are given in figure legends. The reaction was stopped with 85 µl of 0.15 M NaCl, 50 mM Tris-HCl (pH 7.5), 10 mM EDTA and 10 mg/ml tRNA. After phenol extraction, the DNA was precipitated with ethanol, washed with 70% ethanol, dried, dissolved in 1 ml of formamide dye (35), heated for 1 min at 90 °C, rapidly cooled, and applied to denaturing 10% PAG, 40 cm long, with a thickness gradient of 0.15 - 0.45 mm (36). Electrophoresis was run for 75 min at 47 W (2 kV). Prior to exposure, the gel was fixed in 10% acetic acid and dried on glass pretreated with γ-methacryloylpropyloxysilane (LKB).

Molecular modeling studies. An initial canonical B-form DNA structure was generated and minimized for the dodecamer 5'-d(C1A2A3A4A5G6G7T8T9T10T11G12)-3' and its complementary strand, 5'-d(C13A14A15A16A17C18C19T20T21T22T23G24)-3'. This structure was then gradually distorted in order to widen the minor groove at the central nucleotides using a program based on generalized helical parameters as independent variables (37, 38). The bis-netropsin molecule, **bis-Nt-1** (see Figure

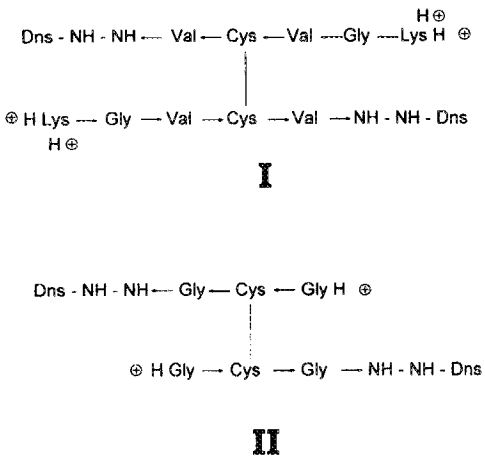


Figure 1: Structure of the cysteine peptides **I** and **II**.

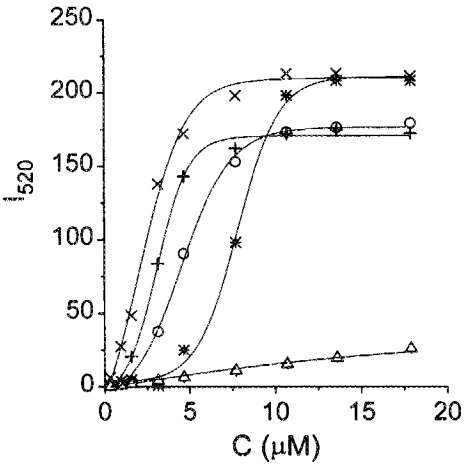


Figure 2: Concentration dependence of peptide **I** fluorescence intensity in the absence of DNA (Δ) and in the presence of poly(dG)•poly(dC) (x), poly(dA)•poly(dT)(O), poly(dI)•poly(dC) (+), poly[d(GC)]•poly[d(GC)] (*). C is peptide concentration. All polynucleotide concentrations are 8×10^{-6} M (base pairs). Experiments were done in 1 mM sodium cacodylate, pH 7.0 in the presence of 0.1 M NaCl at 20 °C.

6), with four positive charges, was constructed in three sections: the two netropsin-like pieces were created by modifying the X-ray structure of a netropsin-duplex complex (39), while the central peptide region was adapted from the NMR structure of triostin A (40). Averages of the Mulliken charges for different molecular conformations were used as partial atomic charges in the netropsin-like terminal portions of the drug, and were calculated using the AM1 semi-empirical molecular orbital method in the MOPAC package (41) in the SYBYL 6.0.3 molecular modeling software package from TRIPOS, Inc (St. Louis, MO.). These charges were in good agreement with those calculated for the true netropsin molecule [S.B. Singh and P.A. Kollman, in preparation]. The partial charges for all non-standard residues in the complex, as well as all force field parameters, are as previously reported (42). The system was rendered electrically neutral by addition of 18 hexa-hydrated Na^+ ions at a distance of 5 Å from the backbone phosphate groups. The electrostatic energy was evaluated using a distance dependent dielectric function, $\varepsilon = R_{ij}$, to mimic bulk solvent effects. The MD simulations, carried out with a time step of 1 fs, and energy minimizations were performed *in vacuo* with the AMBER 4.1 suite of programs (43).The resulting structure were displayed using the MIDASplus molecular graphics display package developed in the Computer Graphics Laboratory, University of California, San Francisco (44).

The netropsin-like ends and the peptide linker were docked by hand into the duplex minor groove as three separate pieces. Model building artifacts were removed by a series of energy minimizations followed by molecular dynamics simulations. Initially, four hydrogen-bonding restraints were imposed between the drug and DNA molecule in order to draw the drug deeper into the minor groove, and energy minimization and MD simulation were repeated. Additional hydrogen bonds were then imposed to find a good starting structure for the longer MD runs. This procedure was repeated several more times, and with each repetition, additional hydrogen bonds were found and used in the next cycle: Ultimately, 14 hydrogen bonds and thirty torsional restraints were included. During 10 ps of MD simulation, the system was heated from 10 K to 300 K, followed by 30 ps of equilibration, and then maintained at that temperature for another 200 ps. The force constants for the distance and torsional hydrogen bonding restraints (except for two loose restrains between the ends of the ligand and the terminal thymines of each strand) were reduced to zero after 15 ps of simulations. A time-averaged structure was produced from the last 70 ps; this was energy minimized with only base-pairing restraints to arrive at an equilibrated and relaxed complex structure.

Results and Discussion

Sequence-specific binding of cysteine peptide dimers to DNA. The chemical structures of the cysteine peptides **I** and **II** are shown in Figure 1. The two peptide chains are linked covalently by a disulfide bond between the two cysteine residues. Each peptide chain carries a fluorescent dansyl chromophore. Two questions were of particular interest for us: (i) does the peptide dimer possess any binding specificity toward nucleotide sequences? (ii) where on DNA is it localized when bound? Fluorescence methods were used to monitor the binding equilibria between the peptide and nucleic acids.

Figure 2 shows typical fluorimetric titration curves obtained for the binding of peptide **I** to various synthetic polydeoxyribonucleotides. One can see that the fluorescence intensity measured at 520 nm is enhanced in the presence of nucleic acids. At low peptide/DNA base pair ratios, the effect decreases in the order poly(dG)•poly(dC) > poly(dI)•poly(dC) > poly[d(AC)]•poly[d(TG)] > poly(dA)•poly(dT) > poly[d(GC)]•poly[d(GC)], reflecting the affinity order shown by the peptide to these synthetic polynucleotides. We found that in the presence of a large molar excess of nucleic acid over peptide, when the free peptide fluorescence was negligible, the intensities calculated per mole of peptide were identical

for binding to poly(dG)•poly(dC) and poly(dA)•poly(dT). Similar results were obtained for binding to poly(dI)•poly(dC) and poly[d(GC)]•poly[d(GC)]. Figure 3 shows typical plots of the fluorescence intensity measured at 520 nm as a function of peptide **I** concentration in the absence and presence of various concentrations of poly(dG)•poly(dC). One can see that data points obtained in the presence of $8x10^{-6}$ M and $3.2x10^{-5}$ M poly(dG)•poly(dC) (base pairs) were almost coincident at low peptide/DNA ratios, indicating that practically all the added peptide was bound to poly(dG)•poly(dC) under these conditions. In the presence of excess poly(dG)•poly(dC), the plot is non-linear, exhibiting a break point at which the fluorescence intensity calculated per mole of bound peptide increases approximately two times relative to the value characteristic of the peptide-DNA complex formed at low extents of binding. Similar results were obtained from binding to naturally occurring DNA's (data not shown). We interpret these results as indicating that there are at least two types of peptide-DNA complexes: the first, formed at very low molar ratios of added peptide to DNA base pairs ($0 < C/BP < 0.03$), when the peptide binds predominantly at isolated sites on DNA and the second, formed at moderate to high peptide/DNA ratios, when binding approaches saturation. Here C is the concentration of the peptide and BP is the concentration of DNA base pairs.

The complex is saturated when two peptide molecules are bound per four base pairs in the case of binding to poly(dG)•poly(dC) and poly[d(GC)]•poly[d(GC)]. Peptide binding to poly(dA)•poly(dT) and poly(dI)•poly(dC) approaches saturation when two peptide molecules are bound per five base pairs. A model which is consistent with these observations, as well as with data obtained in our previous studies (28-30), can be briefly outlined as follows: the peptide binds to DNA as a monomer (β-sheet) and dimer (β-sandwich), each of which occupies 4 to 5 base pairs. Monomers bind to DNA in a non-cooperative manner, while the binding of β-sandwiches occurs cooperatively. These cooperative interactions are assumed to be mediated by direct contacts between nearest neighbor bound dimers.

In the Appendix, we present a statistical mechanical treatment of ligand binding to a polynucleotide lattice for this thermodynamic model. The basic thermodynamic parameters are the binding constants, K_1 and K_2, for binding of the peptide monomer and dimer species to isolated sites on a homogeneous polynucleotide lattice, the dimerization constant K_s, the size of a binding site on DNA, L, and the cooperativity parameter a. The cooperativity parameter represents the dimensionless equilibrium constant for moving a bound dimer from an isolated lattice site to one adjacent to a bound dimer. In order to determine thermodynamic parameters from experimental titration curves, we find it convenient to compare the experimental and theoretically calculated plots of $F = I/I_0 - 1$ versus C/BP (Figure 4), where I is the intensity of fluorescence of the peptide solution at a given peptide concentration in the presence of DNA, I_0 the fluorescence intensity of the free peptide at the same concentration and C/BP is the molar ratio of added peptide to DNA base pairs. If only two bound peptide species are present in solution, then

$$F = ((f_1/f_0-1)*R1 + 2*(f_2/f_0-1)*R2)/(C/BP) \qquad [1]$$

where R1 and R2 are the molar ratios of bound peptide monomers and dimers to DNA base pairs, respectively, f_0 is the fluorescence intensity of the free peptide at a given wavelength calculated per mole of peptide, and f_1 and f_2 are the fluorescence intensities emitted at the same wavelength by bound monomers and dimers, divided by their molar concentrations, respectively. The intensity ratio f_1/f_0 can be determined from measurements of the fluorescence intensity of peptide solutions in the absence of poly(dG)•poly(dC) and in the presence of a large molar excess of poly(dG)•poly(dC). From initial slopes of the plots shown in Figure 3, one can easily calculate that f_1/f_0 is equal to about 13 ± 2. The intensity ratio f_2/f_0 can be determined from fluorescence measurements at high occupancies when binding approaches saturation and bound peptide exists predominantly in the dimer form.

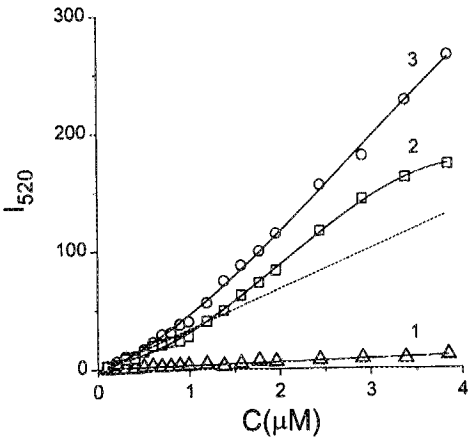


Figure 3: Concentration dependence of peptide **I** fluorescence intensity in the absence of DNA (curve 1) and in the presence of poly(dG)•poly(dC) at a concentration of $8x10^{-6}$ M (curve 2), $3.2x10^{-5}$ M (curve 3). The initial slope of the titration curve, measured in the presence of $3.2x10^{-5}$ M poly(dG)•poly(dC), is indicated by dashed line. Polynucleotide concentrations are in base pairs and conditions are the same as Figure 2.

In this case, one can neglect the contribution of bound monomers to the fluorescence intensity emitted by the peptide-DNA mixture. From data shown in Figure 2 the intensity ratio f_2/f_0 is found to be 36 ± 1 on average for binding to poly(dG)•poly(dC), poly[d(GC)]•poly d(GC)], poly(dA)•poly(dT) and poly(dI)•polyd(C). The theoretical plots of F versus C/BP were calculated from Eq. [1], using experimentally determined values for intensity ratios f_1/f_0 and f_2/f_0 and different values for parameters K_1, a, and $W = K_2K_S$ and for site size L equal to 4 and 5 (Figure 4).

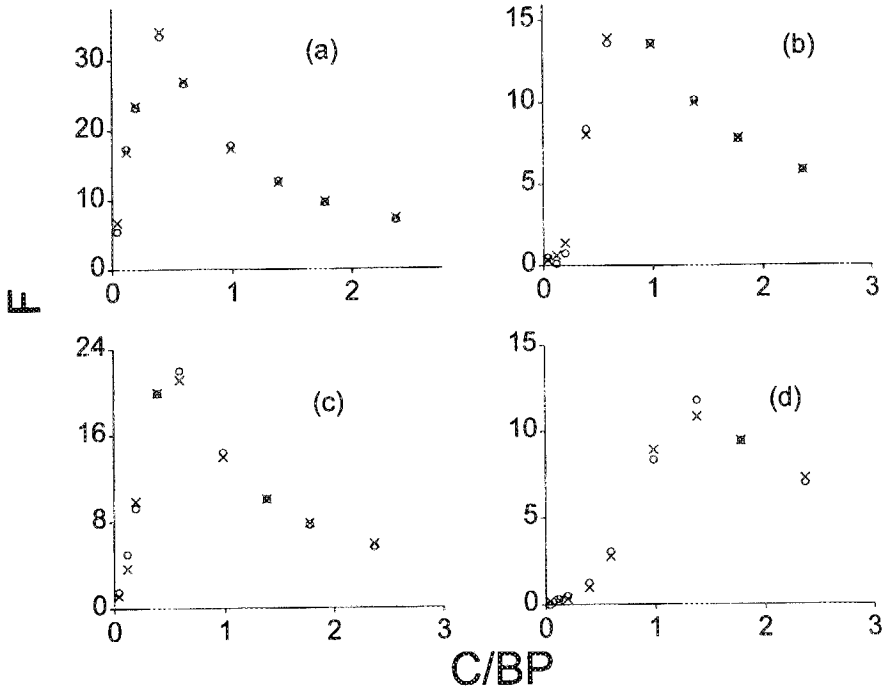


Figure 4: Comparison of the experimental and theoretical plots of F versus C/BP (molar ratio of added peptide to DNA base pairs). Panels a, b, c and d: titrations of poly(dG)•poly(dC), poly(dA)•poly(dT), poly(dI)•poly(dC) and poly[d(GC)]•poly[d(GC)], respectively, by peptide **I**. x, experimental values of F; O , theoretical values of F calculated from Eq. 1 using the best fit values for parameters K_1, a and W. Panel a: $K_1 = 1.7\times10^5$, a = 918, $W = 3.6\times10^9$; panel b: $K_1 = 3.5\times10^3$, a = 2218, $W = 5.9\times10^7$; panel c: $K_1 = 1.0\times10^4$, a = 1005, $W = 5.1\times10^8$; panel d: $K_1 = 2.\times10^3$, a = 470, $W = 6.7\times10^7$. The mean square deviations of the experimental and calculated values of F were equal to 2.6, 1.2, 4.9, and 2.1 for data shown in panels a, b, c and d, respectively.

Thermodynamic parameters were determined by a combination of an iterative and a nonlinear least squares data fitting procedure. The minimum of the mean square deviation between experimental and theoretical values of F calculated for different C/BP values was taken as a criterion of the quality of fit between the experimental and theoretically calculated curves. At first, we neglected the binding of monomers to DNA and calculated concentrations of free and bound peptide using the experimentally determined value of f_2/f_0. The calculated concentrations of the free peptide at different C/BP were used to calculate new values of R1, R2 and F for different values of thermodynamic parameters K_1, a and W. Then a search was carried

Table I

Thermodynamic Parameters for Binding of Cysteine Peptide **I** to Various Synthetic Polydeoxyribonucleotides

Polymer	L	K_1, M^{-1}	W, M^{-2}	a
Poly(dG)•Poly(dC)	4	$(1.2\pm0.6)\times10^5$	$(2.3\pm1.0)\times10^9$	800±200
Poly(dI)•Poly(dC)	5	$(1.0\pm0.4)\times10^4$	$(4.7\pm0.4)\times10^8$	1200±200
Poly(dA)•Poly(dT)	5	$(3.5\pm0.5)\times10^3$	$(3.3\pm1.8)\times10^7$	3100±600
Poly[d(GC)]•Poly[d(GC)]	4	$(1.7\pm0.4)\times10^3$	$(6.7\pm0.5)\times10^6$	800±500
Poly[d(AC)]•Poly[d(TG)]	4	$(8\pm3)\times10^3$	$(1.6\pm0.8)\times10^8$	400±200

Thermodynamic parameters were determined from the experimental titration curves of synthetic DNA's by the peptide. The experimental data on binding of the cysteine peptide to DNA are consistent with a model according to which the peptide binds to DNA both in the monomer (β-sheet) and dimer (β-sandwich) forms. Each bound monomer and each bound sandwich occupies 4 to 5 DNA base pairs. Here K_1 is the binding constant of the monomers to DNA. L is the size of a binding site for the peptide on DNA. $W = K_2K_S$, where K_2 is the binding constant of the dimer to DNA. K_S is the peptide dimerization constant, a is the cooperativity parameter.

out to determine which values of the binding parameters led to a minimum in the mean square deviation between the experimental and calculated values of F. Using the estimated values of the binding parameters K_1, a and W, one can calculate more precisely the free peptide concentration at different C/BP values and then recalculate R1, R2 and F. The process continued until self-consistent values of the binding parameters were obtained.

The average best fit values for thermodynamic parameters K_1, a, L, and W, determined from three independent experiments, are collected in Table I, where one can see that the peptide binding constant to poly(dG)•poly(dC), K_1, is approximately 10 and 30 times greater than that found for binding to poly(dI)•poly(dC) and poly(dA)•poly(dT), respectively. This suggests the importance of specific interaction between the peptide and guanine 2-amino groups exposed in the minor DNA groove. The affinity constant of the peptide monomer to poly[d(GC)]•poly[d(GC)] is approximately 70 times lower than that found for binding to poly(dG)•poly(dC). Since the cysteine peptide contains two identical peptide chains and may assume a symmetrical structure, these results deserve some comments. Evidently, in the complex with poly(dG)•poly(dC), the two chains of the cysteine peptide take up different configurations and recognize 5'-GpG-3' steps on DNA. Interestingly, the estimated values for $W = K_2 K_S$, like the best fit K_1 values, decrease in the following order: poly(dG)•poly(dC) > poly(dI)•poly(dC) > poly(dA)•poly(dT) > poly[d(GC)]•poly[d(GC)], thereby suggesting that specificity determinants of the monomers and dimers (sandwiches) may be virtually identical. The estimated value of the cooperativity parameter varies from one synthetic DNA to another, and is largest for binding to poly(dA)•poly(dT).

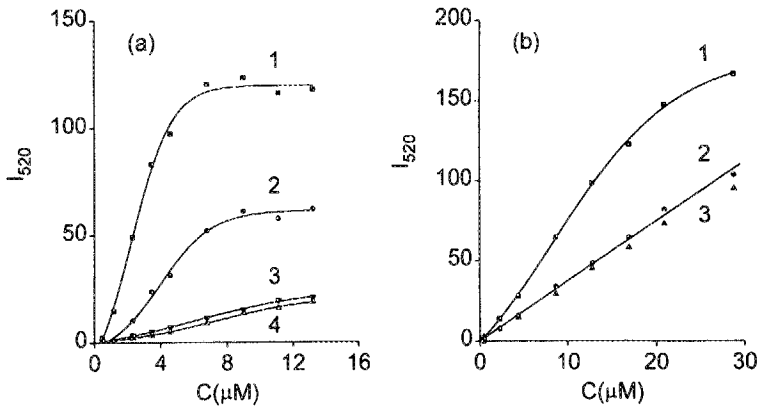
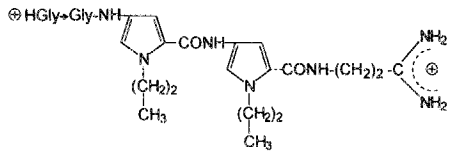
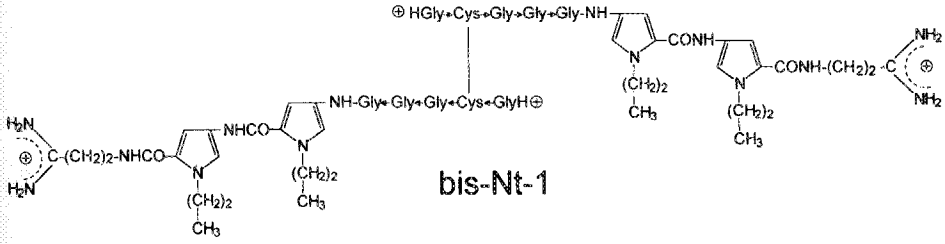


Figure 5: Competition between cysteine peptides **I** and **II** and sibiromycin for binding sites on poly(dG)•poly(dC). Fluorimetric titrations of poly (dG)•poly(dC) complexed with sibiromycin at different molar ratios of bound antibiotic to DNA base pairs by peptides **I** (a) and **II** (b). I_{520} is the fluorescence intensity at 520 nm and C is the peptide concentration. Panel a: curve 1, no sibiromycin; curve 2, sibiromycin/DNA base pair molar ratio is 0.156; curve 3, sibiromycin/DNA base pair molar ratio is 0.435; curve 4, free peptide **I**. Concentration of poly(dG)•poly(dC) was 6.5 x 10^{-6} M (base pairs). Panel b: curve 1, no sibiromycin; curve 2 , sibiromycin/DNA base pair molar ratio is 0.333; curve 3, free peptide **II**. Concentration of poly(dG)•poly(dC) was 7.6 10^{-6} M (base pairs). Same conditions as Figure 2.

In order to investigate in which DNA groove the peptide binds, we have carried out competition experiments using the DNA minor groove binding antibiotics sibiromycin and distamycin A (29). It is known that sibiromycin forms a covalent bond with the guanine 2-amino group in the DNA minor groove (45,46) and that binding does not alter the structure of DNA (47). Fluorimetric titration with peptides **I** and **II** of poly(dG)•poly(dC) alone and poly(dG)•poly(dC) complexed with sibiromycin shows that increasing molar ratios of covalently bound sibiromycin to DNA base pairs lead to a progressive decrease in the fluorescence intensity, which remains proportional to the amount of the peptide bound to the polymer (Figure 5). The amount of sibiromycin covalently bound to the DNA was determined by equilibrium dialysis, as described by Grokhovsky et al. (28). From Figure 5 one can see that no peptide binding can be detected if one sibiromycin molecule is bound per three base pairs. A simple explanation of all these data is that the peptides, like



Gly_2-Nt



bis-Nt-1

Figure 6: Chemical formula of the Gly_2-netropsin analog, **Gly_2-Nt**, and **bis-Nt-1** in which two netropsin-like fragments are linked to a pair of peptides, Gly-Gly-Gly-Cys-Gly-, bridged by a disulfide bond. The direction of N-C^{α} -C' is indicated in the peptide fragments of each compound.

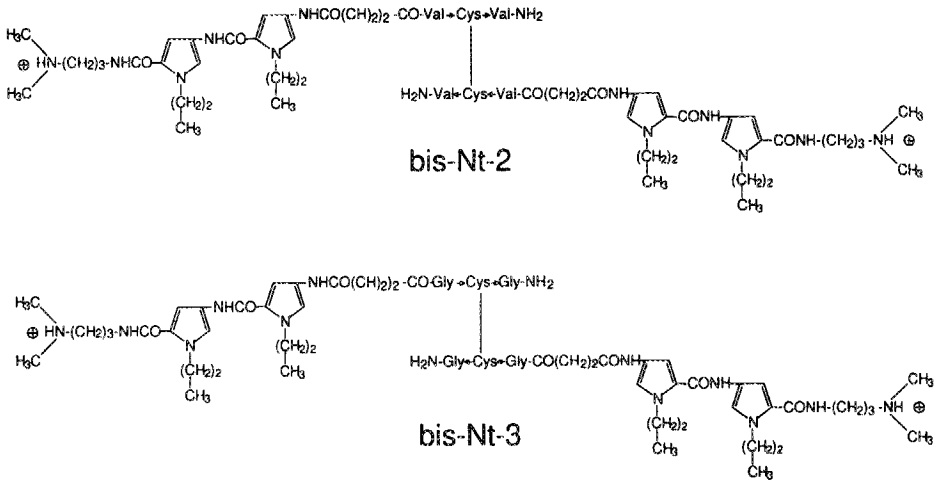


Figure 7: Chemical formulas of bis-netropsins, **bis-Nt-2** and **bis-Nt-3**, in which two netropsin-like fragments are linked to a pair of peptides, -Val-Cys-Val- and -Gly-Cys-Gly-, bridged by a disulfide bond. The direction of N-C^{α} -C' is indicated in the peptide fragments of each compound.

sibiromycin and distamycin A, bind in a sequence-specific manner in the minor groove of DNA.

Bis-netropsin: design, synthesis and interaction with DNA. Our next step was to synthesize netropsin-peptide conjugates containing AT- and GC-specific reaction centers. As AT-specific reaction centers we have chosen Nt-like fragments, while for the GC-specific element we have used cysteine dimeric peptides closely related to peptides **I** and **II**. Two Nt-like fragments were attached to a pair of peptides, -Gly-Cys-Gly- or -Val-Cys-Val -, which were joined with a homologous peptide by means of an S-S-bridge between two cysteine residues. The chemical formulas for a netropsin analog containing two glycine residues at the N-terminus, **Gly_2Nt**, and for three bis-netropsin conjugates, **bis-Nt-1**, **bis-Nt-2** and **bis-Nt-3**, are shown in Figures 6 and 7. **Bis-Nt-1** and **bis-Nt-3** have in the central part of the ligand molecule a pair of peptides Gly-Cys-Gly, bridged by a disulfide bond. In **bis-Nt-2** these peptides are replaced by - Val-Cys-Val- peptides. The three bis-netropsins are also distinguished by the relative orientation of the oligopyrrolecarboxamide and peptide chains. In **bis-Nt-1,** the peptide and N-propylpyrrolecarboxamide groups are linked in such a way that N-C^{α}-C' in each peptide chain coincides with the sequence N-pyrrole C2-pyrrole C4-C'- in the netropsin-like fragment. In **bis-Nt-2** and **3** the peptide and N-propylpyrrolecarboxamide groups are linked in the opposite orientation.

Among three bis-netropsins **bis-Nt-1** exhibits the most interesting DNA-binding properties. Typical CD titration curves of poly(dA)•poly(dT) and poly(dG)•poly(dC) with **bis-Nt-1** are presented in Figure 8. From the position of the breakpoint on the titration curve we can conclude that one bound bis-netropsin molecule occupies approximately 10 to 11 base pairs of poly(dA)•poly(dT) and approximately 12 base pairs of poly(dG)•poly(dC). It is known that netropsin itself and its analog, **Gly_2-Nt**, containing two glycine residues at the N-terminus, occupy 4 to 5 base pairs upon binding (Figure 8). This suggests that the two netropsin-like fragments of the bis-netropsin molecule are implicated in specific interaction with DNA base pairs. Consistent with this interpretation is the fact that the molar dichroism of the **bis-**

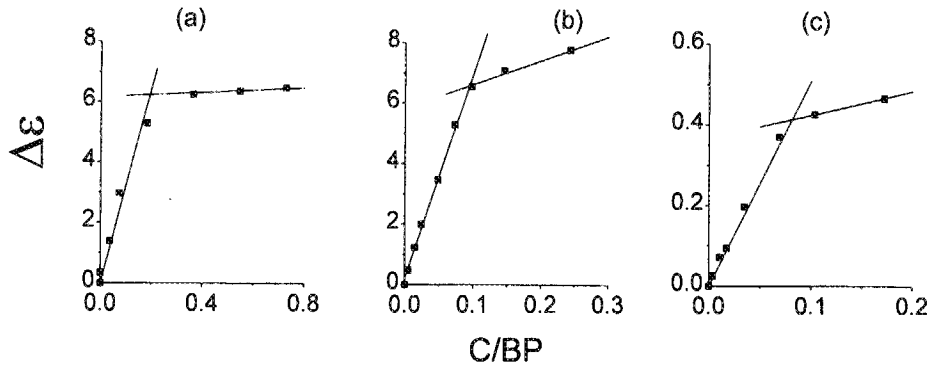


Figure 8: Titration of (a) poly(dA)•poly(dT) (0.1 mM base pairs) by Gly_2-Nt, (b) poly(dA)•poly(dT) (0.62 mM base pairs) by bis-Nt, and (c) poly(dG)•poly(dC) (0.87 mM base pairs) by bis-Nt. Δε is the CD amplitude at 315 nm calculated per mole of base pair. C/BP is the molar ratio of ligand to DNA base pairs. Conditions: (a) 5 mM sodium cacodylate (pH 7), 0.3 M NaCl, 0.5 mM EDTA, (b) and (c) 1 mM sodium cacodylate (pH 7.0), 0.6 M NaCl, 1 mM EDTA. All experiments done at 20 °C.

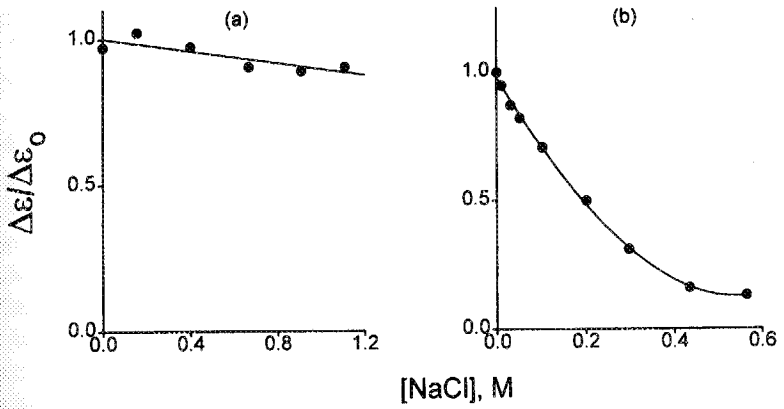


Figure 9: Stability of **bis-Nt-1** complexes with (a) poly(dA)•poly(dT) (0.17 mM base pairs, C/BP = 0.26) and (b) poly(dG)•poly(dC) (0.21 mM base pairs, C/BP = 0.28) as a function of NaCl concentration. $\Delta\varepsilon$ and $\Delta\varepsilon_o$ are the CD amplitude at 315 nm per mole of base pair in the presence and absence of NaCl, respectively.

Nt-1 complex with poly(dA)•poly(dT), 70 ± 3, is approximately twice that found for binding of the netropsin analog **Gly_2-Nt** (33 ± 2). The molar dichroism for **bis-Nt-2** complexed with poly(dA)•poly(dT) is approximately equal to the value obtained for **Gly_2-Nt** whereas the molar dichroism for the complex of **bis-Nt-2** with poly[d(AT)]•poly[d(AT)] is approximately twice as large. Similar results were obtained for binding of **bis-Nt-3** (data not shown), although the molar dichroism of **bis-Nt-3** complexed with poly(dA)•poly(dT) is equal to 60 ± 3.

Another interesting feature of these results is the ability of **bis-Nt-1** to form stable complexes with poly(dG)•poly(dC). This finding is unexpected, because netropsin and its analogues form very weak complexes with poly(dG)•poly(dC), which dissociate in the presence of 0.05 M NaCl (1,3). In Figure 9, plots of the stability of bis-netropsin complexes with poly(dA)•poly(dT) and poly(dG)•poly(dC) as a function of salt concentration are shown. One can see that approximately half of the total amount of bound **bis-Nt-1** dissociates from poly(dG)•poly(dC) in the presence of 0.2 M NaCl. These observations are consistent with the existence of GC-specific reaction centers in the bis-netropsin molecule.

We found that complexes of **bis-Nt-2** and **bis-Nt-3** with poly(dA)•poly(dT) are more sensitive to changes in ionic strength than complexes with **bis-Nt-1**. Approximately half of the total amount of **bis-Nt-3** bound to poly(dA)•poly(dT) in 1 mM sodium cacodylate buffer (pH 7.0) dissociates in the presence of 0.6 M NaCl (31). **Bis-Nt-2** exhibits approximately 10 times lower affinity to poly(dA)•poly(dT) in 1 mM sodium cacodylate buffer as compared to that of **bis-Nt-3**. As illustrated in Figure 10, complexes of **bis-Nt-2** with poly(dA)•poly(dT) dissociate completely in the presence of 0.1 M NaCl. Interestingly, complexes of **bis-Nt-2** with poly[d(AT)]•poly[d(AT)] are more stable (Figure 10). Approximately half of the total amount of bound **bis-Nt-2** dissociates from poly[d(AT)]•poly[d(AT)] in the presence of 0.6 M NaCl. **Bis-Nt-2** interacts more strongly with poly(dG)•poly(dC) than **bis-Nt-3** under the same conditions (data not shown). A characteristic negative CD band is observed at 310 nm upon binding of **bis-Nt-2** to poly(dG)•poly(dC) and poly[d(AC)•poly[d(GT)] which

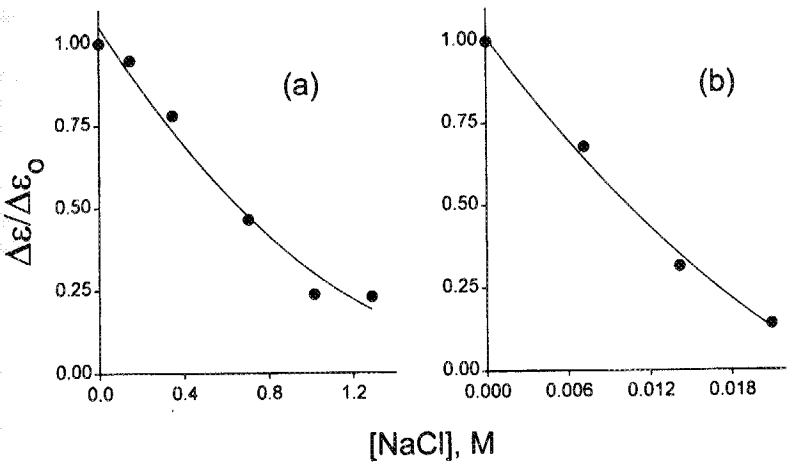


Figure 10: Stability of **bis-Nt-2** complexes with (a) poly[d(A-T)]•poly[d(A-T)] (0.077 mM base pairs, C/BP =0.46) and (b) poly(dA)•poly(dT) (0.113 mM base pairs, C/BP=0.43) as function of NaCl concentration. Notation is the same as in Figure 9. Experiments were done in 1 mM sodium cacodylate buffer (pH 7.0).

can be ascribed to a CD signal induced upon interaction with DNA of the cysteine peptide dimer of the bis-netropsin. Binding of **bis-Nt-2**, in contrast to **bis-Nt-3** (or **bis-Nt-1**), is accompanied by partial precipitation of poly(dG)•poly(dC) and DNA at high bis-netropsin/DNA base pair molar ratios (C/BP>1.0).

In Table II binding constants, determined from experimental adsorption isotherms, for **bis-Nt-1** and **bis-Nt-3** to various naturally occurring and synthetic DNA's are presented. Due to the very high affinity of **bis-Nt-1** for poly(dA)•poly(dT), its binding constants to various naturally occurring and synthetic DNA's were determined by CD methods in the presence of 0.6 M NaCl. In comparison to **bis-Nt-1, bis-Nt-3** exhibits approximately 100 times lower affinity to poly(dA)•poly(dT) in the presence of 0.6 M NaCl (Table II). Binding constants of **bis-Nt-3** to various naturally-occurring and synthetic DNA's were measured in the absence and presence of 0.6 M NaCl (see Table II).

The presence of reaction centers specific for both AT and GC base pairs can be revealed in the characteristic form of the curve describing the dependence of lnK on $\ln X_{AT}$, where X_{AT} is the fraction of AT pairs in naturally occurring DNA and K is the binding constant. It is easy to show that if a ligand carries n reaction centers capable of interacting preferentially with AT base pairs and does not contain functional groups interacting preferentially with GC pairs, then the theoretical depen-

Table II

Binding Constants of **Bis-Nt-1** and **Bis-Nt-3** to Various Naturally Occurring and Synthetic DNA's

DNA	X_{AT}	**Bis-Nt-3** K, M^{-1} [a]	**Bis-Nt-3** $K\cdot, M^{-1}$ [b]	**Bis-Nt-1** $K,\cdot M^{-1}$ [b]
Poly(dA)•Poly(dT)	1.00	$(200\pm20)\times10^5$	$(3.6\pm0.5)\times10^4$	$(40\pm5)\times10^5$
Poly[d(AT)]•Poly[d(AT)]	1.00	$(3.0\pm0.5)\times10^5$	$(1.0\pm0.2)\times10^4$	$(8\pm2)\times10^5$
Clostridium Perfringence DNA	0.72	$(30\pm3.0)\times10^5$	$(1.2\pm0.2)\times10^4$	$(6.0\pm1.5)\times10^5$
Calf Thymus DNA	0.58	$(8.0\pm1.0)\times10^5$	$(0.7\pm0.2)\times10^4$	$(2.0\pm0.5)\times10^5$
Escherichia Coli DNA	0.50	$(5.0\pm0.7)\times10^5$	$(0.5\pm0.1)\times10^4$	$(1.2\pm0.6\times10^5$
Thermus Species DNA	0.36	$(1.8\pm0.3)\times10^5$	$(0.9\pm0.2)\times10^4$	$(1.0\pm0.5)\times10^5$
Micrococus Lysodeikticus DNA	0.28	$(1.5\pm0.3)\times10^5$	$(1.3\pm0.2)\times10^4$	$(0.8\pm0.4)\times10^5$
Poly(dG)•Poly(dC)	0.00	$(2.5\pm0.7)\times10^5$	not determined	$(1.5\pm0.5)\times10^5$

Binding constants were determined from the experimentally measured CD titration curves of various naturally occurring and synthetic DNA's by synthetic ligands. Experiments were done at 20 °C either in 1 mM sodium cacodylate buffer (pH 7.0) in the presence of 0.1 mM EDTA (a) or in 10 mM sodium cacodylate buffer (pH 7.0) in the presence of 1 mM EDTA and 0.6 M NaCl (b). Bis-netropsins in free solution were found to be optically active with a molar dichroism at 315 nm of 0.65 ± 0.05 and 3.5 ± 0.5 for **bis-Nt-1** and **bis-Nt-3**, respectively. The molar dichroism of complexes between ligand and various DNA's was determined in a separate series of experiments performed in the presence of a large molar excess of nucleic acid over synthetic ligand. In the case of **bis-Nt-1** complexes with calf thymus DNA, *Escherichia coli* DNA, *Thermus species* DNA, *Clostridium perfringence* DNA, *Micrococus lysodeikticus* DNA, poly(dA)•poly (dT), poly [d(AT)]•poly[d(AT)], poly(dG)•poly(dC), the molar dichroism values at 315 nm are 60 ± 2, 60 ± 2, 40 ± 2, 80 ± 5, 10 ± 3, 70 ± 3, 83 ± 5, 9 ± 3 , respectively. In the case of complexes of **bis-Nt-3** with the same set of naturally-occurring and synthetic DNA's the molar dichroism values are 40 ±2, 40 ± 3, 25 ± 5, 50 ± 3, 15 ± 5, 60 ± 2, 50 ± 2, 8 ± 3 , respectively. The binding data were analyzed as Scatchard plots, r/m versus r, where r is the molar ratio of bound ligand to DNA base pair and m is the concentration of free ligand. The binding constants were determined from the intercept of the initial linear part of Scatchard isotherm on the vertical (r/m) axis (for details see ref. 31).

dence of lnK on $\ln X_{AT}$ can be calculated according to Eq. [2] :

$$\ln K = \ln K_o + n \ln S + n \ln X_{AT} \qquad [2]$$

where K_o is a constant characterizing the contribution of non-specific interactions between ligand and DNA to the overall binding constant, S is the stability constant of a bond between a single ligand reaction center and an AT base pair, and X_{AT} is the fraction of AT base pairs in the DNA. From Eq. [2] one can conclude that the total number of ligand reaction centers, n, can be determined from the slopes of the experimental curves of lnK versus $\ln X_{AT}$. Figure 11 shows the theoretical plots of lnK versus $\ln X_{AT}$ calculated from Eq. [2] for n changing from 1 to 6 and for the same values of parameters S and K_o. Each plot is a straight line with a slope equal to n. The data points obtained from the binding of the netropsin analog **Gly_2-Nt** to various DNA's all lie on a straight line with a slope equal to 3 (Figure 11). This indicates that **Gly_2-Nt** carries 3 AT-specific reaction centers, which can be associated with its carboxamide NH groups. In contrast, the data points obtained for binding of **bis-Nt-1** to various naturally occurring and synthetic DNA's deviate significantly from a straight line (Figure 11).

We have shown earlier (31,48) that, if the ligand contains n_1 AT- and n_2 GC-specific reaction centers, then in the limits when $X_{AT} \rightarrow$ and $X_{GC} \rightarrow 1$,

$$n_1 = \lim_{X_{AT} \rightarrow 1} \frac{d \ln K}{d \ln X_{AT}} \qquad [3]$$

and

$$n_2 = \lim_{X_{GC} \rightarrow 1} \frac{d \ln K}{d \ln X_{GC}} \qquad [4]$$

It can be easily seen from Figure 11 that the initial slope of the plot of lnK versus $\ln X_{AT}$, when $X_{AT} \rightarrow 1$, is equal to 6 ±1, whereas the plot of lnK versus $\ln X_{GC}$ has a slope equal to 2 ±1, when $X_{GC} \rightarrow 1$. Thus, **bis-Nt-1** behaves as a two-component ligand with six AT- and two GC-specific reaction centers. In our previous work, it had been shown that the number of amido groups in the netropsin-like fragments correlates with the number of AT-specific centers determined from initial slopes of plots lnK versus $\ln X_{AT}$ (7, 31). Since the N-propylpyrrole carboxamide fragments of the bis-netropsin molecule serve as AT-specific reaction centers, we suggest that a pair of peptide fragments in the central part of the bis-netropsin molecule carry two GC-specific reaction centers. The nonlinear character of the dependence of lnK on $\ln X_{AT}$ and of lnK on $\ln X_{GC}$ obtained for complexes of DNA with **bis-Nt-1** clearly indicates that it contains not only AT-specific reaction centers, but also GC-specific reaction centers. Nonlinear plots of lnK versus $\ln X_{AT}$ and of lnK versus $\ln X_{GC}$ were also obtained for binding of **bis-Nt-2** and **bis-Nt-3** (data not shown). For binding of

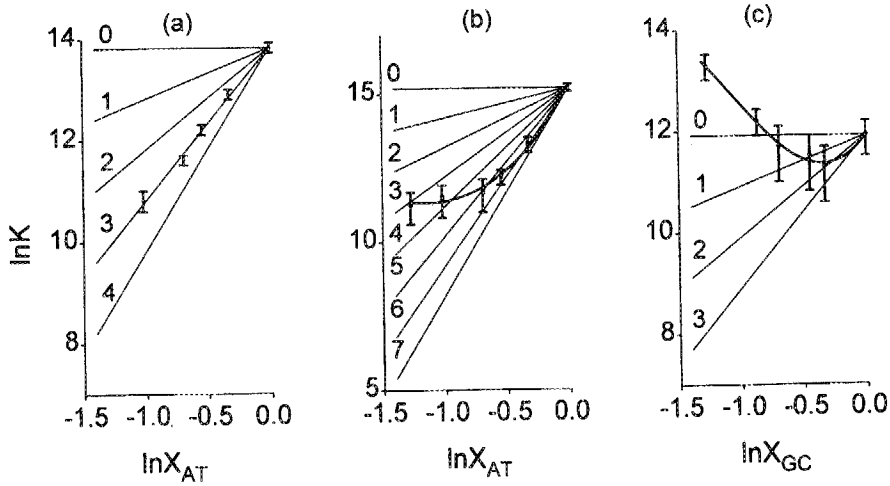


Figure 11: Plots of lnK versus $\ln X_{AT}$ and lnK versus $\ln X_{GC}$ obtained for the binding of (a) **Gly_2-Nt** and (b,c) **bis-Nt-1** to various naturally-occurring and synthetic DNA's. Straight lines represent theoretical dependence of lnK on $\ln X_{AT}$ (or $\ln X_{GC}$) calculated for binding of a one-component ligand with the indicated number of AT- (or GC)- specific reaction centers in a ligand. Binding constants for each ligand were measured under same conditions as Figure 8.

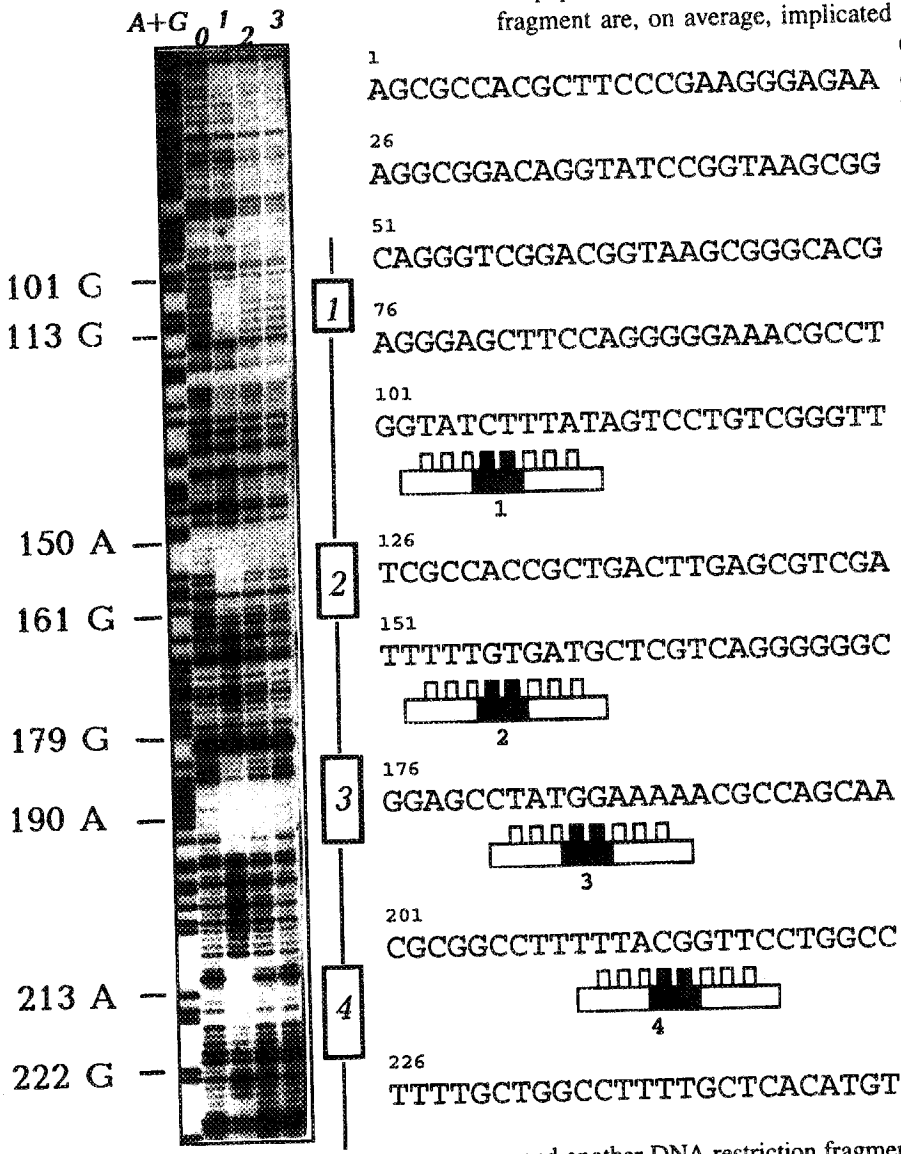


Figure 12: Autoradiogram of the DNase I protection patterns of the DNA fragment A in the presence of **bis-Nt-1**. The ^{32}P end-labeled DNA fragment A (~100 Bq) and unlabelled carrier DNA (10 µl) at 0.01 mM (base pairs) concentration were incubated with **bis-Nt-1** at 20 °C for 10 min in 10 mM Tris HCl buffer (pH 7.6) in the presence of 0.1 M NaCl and then 4 µl of enzyme solution (5 µg/ml) in 10 mM Tris HCl buffer (pH 7.6) in the presence of 0.1 M NaCl and 5 mM $MgCl_2$ was added. The mixtures were kept at 0 °C for 1, 4, 2.5 and 1.5 min. The cleavage products generated by DNase I are shown in lanes 0 to 3, respectively. Lane 0, control, cleavage of the free DNA. Concentrations of **bis-Nt-1** were as follows: lane 1, $2.6xx10^{-6}$ M ; lane 2, $8.7x10^{-7}$ M and lane 3, $2.9x10^{-7}$ M. The protected zones and preferred interaction sites for **bis-Nt-1** on DNA fragment A are indicated. Each protected zone extends over 13 - 14 nucleotides. The center of gravity of the interaction site for **bis-Nt-1** is shifted relative to the center of gravity of the corresponding protected zone on the autoradiogram by 3 nucleotides in the 5' direction.

bis-Nt-3, the initial slopes of plots of lnK versus lnX_{AT} and of lnK versus lnX_{GC} were approximately equal to 6 ± 1 and 2 ± 1, respectively, in close similarity to the results obtained for binding of **bis-Nt-1**. In the case of binding of **bis-Nt-2,** the initial slopes n_1 and n_2 were equal to 3 ± 1 and 2 ± 1, respectively. This suggests that the peptides in the central part of the **bis-Nt-2** molecule and only one netropsin-like fragment are, on average, implicated in specific interaction with DNA base pairs. Consistent with this interpretation are our observations that the size of a binding site for **bis-Nt-2** on poly(dA)•poly(dT) extends over 5 to 6 base pairs only.

Preferential binding sites for bis-netropsin on DNA fragments with known base pair sequences. It was of interest to examine which base pair sequences would be protected from DNase I cleavage in the bis-netropsin complexes with a DNA restriction fragment. Figure 12 presents an autoradiograph of the DNase I protection patterns observed in the presence of various concentrations of **bis-Nt-1,** along with the nucleotide sequence of the restriction fragment being used. As is clear from Figure 12, four distinct regions protected by the bis-netropsin molecule from hydrolysis by DNase I are observed. The strongest affinity binding site for which some protection is observed in the presence of as little as 0.3 µM **bis-Nt-1** has the nucleotide sequence 5'-CTATGGAAAA-3'. Other binding sites are occupied in the presence of higher bis-netropsin concentrations. DNase I footprinting diagrams generated in the presence of **bis-Nt-2** and **bis-Nt-3** will be published elsewhere. Preliminary results show that specificity determinants of **bis-Nt-1** and **bis-Nt-3** are very similar, if not identical.

Figure 13 shows DNase I footprinting diagrams generated in the presence of **bis-Nt-1** and another DNA restriction fragment which is part of a human urokinase gene (34). One can see that there are three clearly defined protection zones, each of which can be associated with an interaction site for bis-netropsin. The strongest affinity binding site for **bis-Nt-1** has a nucleotide sequence 5'-TTTTCCAAAG-3', as revealed by the concentration dependence of **bis-Nt-1** mediated cleavage protection. This sequence contains blocks of four T's and three A's separated by a CpC step. The similarity in the nucleotide sequences found in the two strongest affinity sites for bis-Nt and literature data showing that netropsin itself binds preferentially to runs of four T's or A's on DNA provides a basis for suggesting that the strongest affinity sites observed by us for **bis-Nt-1** binding represent variants of the sequence 5'-TTTTCCAAAA-3' which is actually recognized by **bis-Nt-1** on DNA. This is supported by direct measurements of binding affinity of **bis-Nt-1** to various DNA duplexes with defined nucleotide sequences (49).

Molecular modeling of bis-netropsin-DNA complexes. The goal of our molecular mechanics and dynamics calculations is to provide an overall picture of the complex formed between **bis-Nt-1** and DNA at the molecular level. As can be seen from the structure of this ligand bound to a short DNA duplex, illustrated in Figure 14, the

linear combination of two netropsin-like groups, separated by the double-stranded oligopeptide, follows the helical curvature of the duplex minor groove and this complementarity in shape provides the basis for a tight complex. Each netropsin-like end of the ligand binds to the A-T-rich region of the duplex in a manner generally similar to the netropsin-DNA complex (7,38). Furthermore, the amide protons at the ends of the ligand form bifurcated hydrogen bonds with both the O2 of thymines and the N3 of adenines in the neighboring base pairs. The same type of bifurcated hydrogen bonding has been observed in the X-ray structures of complexes of both netropsin and distamycin with DNA (8,9,39). The beta-sheet character of the central peptide portion of the molecule remains largely intact throughout the time course of the MD simulations. The peptide fragments may form hydrogen bonds connecting the backbone CO groups with the guanine 2 amino groups (50). This recognition mechanism is observed in the crystalline complex of quinoxaline antibiotic triostin A with a self-complementary hexanucleotide 5' - CGATCG-3' (23), in which the cyclic peptide of the antibiotic is accommodated in the minor groove. The NH and CO groups of its backbone form hydrogen bonds with two GC pairs. Such complexation is accompanied by significant conformational alterations in the DNA, resulting in a narrowing of the AT region of the minor groove and a widening of the GC region of the minor groove. This is in line with our earlier model for the insertion of two antiparallel peptide segments into the DNA minor groove (50). More detailed calculations to compare structures of bis-netropsin complexes with different DNA sequences are in progress.

The crystalline complex of transcription activator TBP with a DNA fragment exhibits even greater structural changes than those observed in the triostin-DNA complex (51,52). These changes are probably of significance for initiation of transcription. Although in the TBP-DNA complex the backbone NH and CO groups of the two β-stranded segments of TBP are not implicated in hydrogen bonding with DNA bases, the general plan of the structure of DNA complexes with triostin A and TBP appear similar. In the two complexes, the N--C^{a}--C' sequence in the peptide segment coincides with the C3'--C5' direction in the nearest polynucleotide strand; the peptide chains are inserted deeply in the DNA minor groove with increased width and do not lie on the outer surface of DNA in the B-conformation as assumed in the model of Church et al. (53). One may speculate that formation of a TBP-DNA complex with a geometry similar to that of the triostin-DNA complex may serve as an intermediate step in generating the final structure of TBP-DNA complex. Our experimental data suggest that the two antiparallel β-stranded chains bridged by a S-S bond can be inserted in the DNA minor groove. This conclusion follows from the fact that peptide chains attached

- TGAGT CTCCC TGGCA GGAAT
- CTGTT TCCCA CTGTG GGTCA
- GCAGC ACACA CAGCA TTTTG
- GTGGT GACTT CAGAG CCGTA
- GTAGT GGGGC TGCTG ACACT
- CCCGG TGGGA AATCA GCTTC
- ACAAC AGTCA TTTTC AGCTG
- CTCCG GATAG AGATA GTCGG
- TAGAA TTCTC TTTTC CAAAG
- CCAGT GATCT CACAG CTTGT
- GCCAA ACTGG GGATC GTTAT
- ACATC G

Figure 13: DNase I cleavage protection patterns generated by DNase I on the DNA fragment B in the presence of **bis-Nt-1**. A+G are purine sequence markers. The ^{32}P end-labeled DNA fragment B (~100 Bq) and unlabelled λ phage DNA (5 μl) at 0.06 mM concentration (base pairs) were incubated with bis-netropsin at 20 °C for 10 min in 10 mM Tris HCl buffer (pH 7.6) in the presence of 0.2 M NaCl and then 5 μl of the enzyme solution (5 μg/ml) in 10 mM Tris HCl buffer (pH 7.6) in the presence of 0.1 M NaCl and 5 mM $MnCl_2$ was added. The mixtures were kept at 0 °C for 2, 5, 3, 2 and 5 min, and cleavage products generated by DNase I are shown in lanes 0 to 4, respectively. Lane 0, control, cleavage of the free DNA. Concentrations of **bis-Nt-1** were as follows: lane 1, $1x10^{-6}$ M; lane 2, $2x10^{-6}$ M; lane 3, $4x10^{-6}$ M, lane 4, $8x10^{-6}$ M. The protected zones and the preferred interaction site for **bis-Nt-1** on fragment B are indicated.

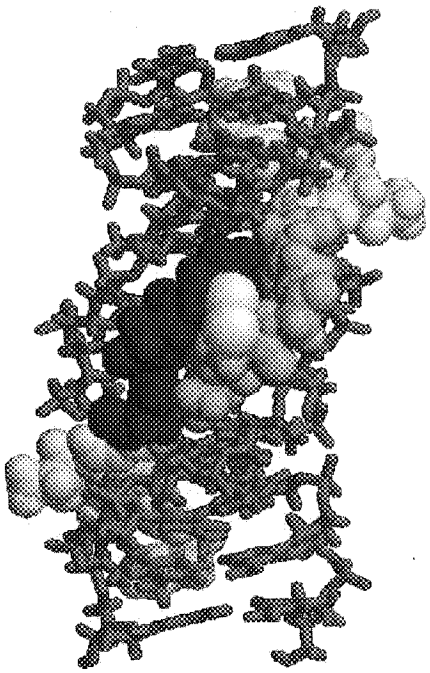


Figure 14: Model of -DNA complex viewed from the side of the minor groove. **Bis-Nt-1** is displayed in space-filling mode, with the two netropsin fragments lightly shaded and each of the two peptide chains shaded in a different manner. The central, light spheres represent the two S atoms. In the complex, the sequence N-C^{α}-C' in each peptide fragment of **bis-Nt-1** coincides with the direction C3'-C5' in the adjacent polynucleotide chain. The netropsin-like fragments form bifurcated hydrogen bonds with the two clusters of A-T base pairs, as observed in the crystalline structures of netropsin-DNA complexes (Kopka et al., 1985; Tarbenero et al., 1993). See text for details of calculations.

to netropsin-like fragments via short linkers can be implicated in interaction with GC base pairs upon binding of the netropsin-like fragments of bis-netropsin to DNA. Furthermore, the cysteine peptide dimer itself (without the netropsin-like fragments) competes with the minor groove binding antibiotics sibiromycin and distamycin A. Our observations that **bis-Nt-3** exhibits the same specificity, but lower affinity for binding to DNA than **bis-Nt-1** can be attributed to the presence of the dimethyl amino groups at the ends of the ligand. They are replaced with amino groups at the ends of **bis-Nt-1**. In the crystalline netropsin-DNA complex the amidine and propionamidine ends of the antibiotic molecule are implicated in hydrogen bond formation with AT-base pairs (8,9). These interactions, including electrostatic component and hydrogen bonds, provide additional stabilization of the antibiotic-DNA complex .

Summary: The use of a two-stranded peptide motif in combination with the synthesis of netropsin-peptide conjugates offers a new approach to the design of compounds capable of recognizing different nucleotide sequences in the minor groove of DNA. The recognition mechanism implicating the backbones of the two antiparallel peptide chains in interaction with functional groups of base pairs exposed in the minor groove appears not to be restricted solely to DNA complexes with antibiotics of the triostin class. It may well be a widespread phenomenon in protein-nucleic acid interactions.

Acknowledgments

It is a pleasure to thank Dr. N. Ulyanov for providing the distorted duplex structure used to initiate our modeling studies, and Jennifer L. Miller and Dr. Suresh B. Singh for helpful discussions on the use of AMBER 4.1. We gratefully acknowledge support for this work from a Fogarty International Research Collaboration Award, NIH (grant 5R03 TW00145), the National Institute of General Medical Sciences, NIH, (grant GM 51650) and from the Russian Foundation for Fundamental Studies. We also acknowledge the Computer Graphics Laboratory, University of California, San Francisco, supported by grant RR01081 from the Division of Research Resources, National Institutes of Health, U.S Department of Health and Human Services.

Appendix

Ligand binding in monomeric and self-associated sandwich forms to a polynucleotide lattice

We consider here the case in which a ligand can bind to a polynucleotide lattice as a monomer or a sandwich composed of two ligand molecules. Each monomer and sandwich occupies L consecutive base pairs upon binding. Let K_1 and K_2 be the equilibrium binding constants of monomer and sandwich forms, respectively, to an isolated site on the polynucleotide lattice. In a bound sandwich, the two ligand molecules are nonequivalent: one of them is implicated in specific interactions with functional groups on DNA bases, whereas the other interacts with the sugar-phosphate backbone. Binding of the monomers is assumed to be non-cooperative; in contrast, bound dimers are allowed to interact with nearest neighbor bound sandwiches. Cooperativity can be quantitated via the quantity a, a dimensionless equilibrium constant for transfer of a bound ligand from an isolated lattice site to a site adjacent to another bound ligand. A bound sandwich can interact with either one or two nearest neighbor bound sandwiches, characterized by binding constants aK_2 and a^2K_2, respectively.

It is well known that the average value of any thermodynamic parameter or function, including the average occupancy of the polymer by ligand in monomer or sandwich forms, can be calculated from the Gibbs grand canonical ensemble parti-

tion function, Ξ_N. In order to evaluate Ξ_N, we apply a recursive method, described in detail elsewhere (Gursky & Zasedatelev, 1984). The essence of this approach is to express the grand partition function for the system containing a polymer chain of N residues as a linear combination of the partition functions for systems containing polymers with a smaller number of residues. It can easily be shown, for the thermodynamic system under study, that

$$\Xi_N = \Xi_{N-1} + K_1 m \Xi_{N-L} + K_2 K_S m^2 \Xi_{N-L-1} + a K_2 K_S m^2 (\Xi_{N-L} - \Xi_{N-L-1}) \qquad [A1]$$

where K_S is the dimerization constant and m is the free ligand concentration.

Equation [A1] has a clear physical meaning. The partition function Ξ_N accounts for all the states of the system, including the states in which the Nth polymer residue is either free or occupied by a ligand. The states corresponding to the situation when the Nth residue is free are all present in Ξ_{N-1}. Thus all of the states in which the Nth residue is occupied by a ligand are described by the difference Ξ_N - Ξ_{N-1}. From eq. [A1] it follows that this difference can be represented as a sum of statistical weights of two events corresponding to cooperative and non-cooperative ligand binding to a site containing the Nth polymer residue. Cooperative binding is possible only when the (N-2L), (N-2L+1)th....(N-L)th residues are occupied by a sandwich, whereas the (N-L+1)th, (N-L+2)th.....Nth residues are all free and can be occupied by another sandwich. The term $aK_2K_sm^2(\Xi_{N-L}-\Xi_{N-L-1})$ represents the sum of statistical weights describing cooperative ligand binding in a self-associated form (sandwich) at the end of the polynucleotide lattice, where as the term $K_1m\Xi_{N-L}$ is the sum of statistical weights for ligand binding in the monomer form to a site containing the Nth polymer residue.

As N approaches infinity, Ξ_N is asymptotic to X_1^N, X_1 being a constant. The algebraic equation which X_1 must satisfy can be obtained by dividing both sides of eq.[A1] by Ξ_N and then going to the limit as N approaches infinity:

$$X^{L+1} - X^L - K_1 m X - a K_2 K_S m^2 (X-1) - K_2 K_S m^2 = 0 \qquad [A2]$$

The average quantity of bound ligand per mole of polymer residue, r, is given by the following relation:

$$r = \frac{\partial \ln X_1}{\partial \ln m} \qquad [A3]$$

where X_1 is the largest root of eq. [A2]. Combining eqs. [A2] and [A3], one can obtain eqs. [A4] and [A5]:

$$r = \frac{K_1 m + 2K_2 K_S m^2 \left(\dfrac{aX_1 - a + 1}{X_1} \right)}{(L+1)X_1^L - LX_1^{L-1} - K_1 m (1 + aK_2 K_S m^2)} \qquad [A4]$$

$$m = C - r[DNA] \qquad [A5]$$

where C is the total ligand concentration and [DNA] is the nucleic acid concentration expressed in moles of base pairs.

References and Footnotes

1. Ch. Zimmer and U. Wahnert, *Prog. Biophys. Mol. Biol. 47*, 31-112 (1986).
2. P.B. Dervan, *Science 232*, 464-471 (1986).

3. A.S. Zasedatelev, G.V. Gursky, Ch. Zimmer and H.Thrum, *Mol. Biol. Rep. 1*, 337-342 (1974).
4. A.S. Zasedatelev, A.L. Zhuze, Ch. Zimmer, S.L. Grokhovsky, V.G. Tumanyan, G.V. Gursky and B.P. Gottikh, *Dokl. Akad. Nauk USSR 231*, 1006-1009. (1976).
5. G.V. Gursky, V.G. Tumanyan, A.S. Zasedatelev, A.L. Zhuze, S.L. Grokhovsky and B.P. Gottikh, In *Nucleic Acid-Protein Recognition*, (Vogel, H.,Ed.) Acad. Press, New York, p. 189-217 (1977).
6. A.S. Krylov, S.L. Grokhovsky, A.S. Zasedatelev, A.L. Zhuze, G.V.Gursky and B.P. Gottikh, *Nucleic Acids Res. 6*, 289-304 (1979).
7. G.V. Gursky, A.S. Zasedatelev, A.L. Zhuze, A.A. Khorlin, S.L. Grokhovsky, S.A. Streltsov, A.N. Surovaya, S.M. Nikitin, A.S. Krylov, V.O. Retchinsky, M.V. Mikhailov, R.Sh. Beabealaschvilli and B.P. Gottikh, *Cold Spring Harbor Symp. Quant. Biol. 47*, 367-378 (1983).
8. M.L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R.E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376-1380 (1985).
9. M. Coll, C.A. Frederick, A.H.-J Wang and A. Rich, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385-8389 (1987).
10. R.E. Klevit, D.E. Wemmer and B.R. Reid, *Biochemistry. 25*, 3296-3303 (1986).
11. J.W. Lown, K. Krowicki, U.G. Bhat, B. Ward, J.C. Dabrowiak, *Biochemistry 25*, 7408-7416 (1987).
12. K. Kissinger, K. Krowicki, J.C. Dabrowiak and J.W. Lown, *Biochemistry 26*, 5590- 5595 (1987).
13. W.S. Wade and P.B. Dervan, *J. Am. Chem. Soc. 109*, 1574-1575 (1987).
14. G. Burkhardt, H. Votavova, J. Spooner, G. Luckand and Ch. Zimmer, *J. Biomol. Struct. Dyn. 2*, 721-736 (1985).
15. J.G. Pelton and D.E. Wemmer, *Proc. Natl. Acad. Sci. U.S.A. 86*, 5723-5727 (1989).
16. J.G. Pelton and D.E. Wemmer, *J. Am. Chem. Soc. 112*, 1393-1399 (1990).
17. X. Chen, B. Ramakrishnan, S.T. Rao and M. Sudaralingam, *Struct. Biol. 1*, 169-175 (1994).
18. W.S. Wade, M. Mrksich and P.B. Dervan, *J. Am. Chem. Soc.114*, 8783-8792 (1992).
19. M. Mrksich, W.S. Wade, T.J. Dwyer, B.H. Geierstanger, D.E. Wemmer and P.B. Dervan, *Proc. Natl. Acad. Sci. U.S.A. 89*, 7586-7590 (992)
20. T. Dwyer, B. H. Geierstanger, Y. Bathini, J.W. Lown, D.E. Wemmer, *J. Am. Chem. Soc. 114*, 5911-5919 (1992).
21. B.H. Geierstanger, T. Dwyer, Y. Bathini, J.W. Lown, D.E. Wemmer, . *J. Am. Chem. Soc. 115*, 4474-4482 (1993).
22. B.H. Geierstanger, M. Mrksich, P.B. Dervan and D.E. Wemmer, *Science 266*, 646- 650 (1994).
23. M.J. Waring, and K.R. Fox, In *Molecular Aspects of Anticancer Drug Action* (Neidle, S. and Waring, M.J., Eds) pp 127-156 Verlag Chemie (1983).
24. A.M.-J Wang, G. J. Ughetto, T.Quigley, T. Hakoshima, G.A. van der Marel, J.H. van Boom and A. Rich, *Science. 225*, 1115-1123 (1984).
25. K.J. Addess, J.C. Sinsheimer and J. Feigen, *Biochemistry 23*, 2496-2508 (1993).
26. S.A Streltsov, A.A. Khorlin, A.N Surovaya, G.V. Gursky, A.S Zasedatelev, A.L. Zhuze and B.P. Gottikh, *Biofizika (USSR) 25*, 929-941 (1980).
27. Yu.Yu Vengerov, T.E. Semenov, A.N Surovaya, N.Yu. Sidorova, S.A. Streltsov, A.A. Khorlin, A.L. Zhuze and G.V. Gursky, *J. Biomol. Struct. Dyn. 6*, 311-330 1988.
28. S.L. Grokhovsky, A.N. Surovaya, N.Yu. Sidorova, H. Votavova, J. Sponar, I. Frich and G.V.Gursky, *Mol..Biol. 22*, 1315-1333 (1988).
29. N.Yu. Sidorova, A.N.Surovaya, V.A. Nikolaev, A.L Zhuze and G.V. Gursky, *Mol. Biol. 25*, 706-717 (1991).
30. T.A Leinsoo, V.A. Nikolayev, S.L.Grokhovsky, S.A. Streltsov, A.S. Zasedatelev, A.L. Zhuze and G.V. Gursky, *Mol..Biol. 22*, 159-179 (1988).
31. T.A. Leinsoo, V.A. Nikolaev, S.L.Grokhovsky, A.N. Surovaya, N.Yu. Sidorova, S.A. Streltsov, A.S. Zasedatelev, A.L Zhuze and G.V Gursky, *Mol. Biol. 23*, 1616-1637 (1989).
32. Yu. A. Ovchinnikov, S.O. Guryev, A.S.Krayev, G.S Monastyrskaya, K.G. Skryabin, E.D. Sverdlov, V.M. Zhakaryev and A.A. Bayev, *Gene 6*, 235-249 (1979).
33. J.Sambrook, E.F Fritsch and T. Maniatis, *Molecular Cloning. A Laboratory Manual*, 2nd ed. Cold Spring Harbour Laboratory, Cold Spring Harbor, New York (1989)
34. P. Jacobs, A. Cravador, R. Loriau, F. Brockly, B. Colau, P. Chuchana, A. Von Elsen, A. Herzog and A. Bollen, *DNA 4*, 139-146 (1985).
35. D.J. Galas and A. Schmitz, *Nucleic Acids Res. 5*, 3157-3170 (1976).
36. A.M. Maxam, and W.Gilbert, *Methods in Enzymology 65*, 499-560 (1980).
37. N.B. Ulyanov, A.A. Gorin and V.B. Zhurkin, In *Proceedings of the International Conference on Supercomputing 89: Supercomputer Applications*, (L.P. Kartashev and S.L. Kartashev; Eds) pp 368-370, International Supercomputing Institute, Inc., St. Petersburg, FL (1989).
38. V.B. Zhurkin, A.A Gorin, A.A. Charakhchyan and N.B. Ulyanov, In *Theoretical Biochemistry and Molecular Biophysics*, (Beveridge, D.L. and Lavery, R., Eds) pp 409-429, Adenine Press, New York 1990).
39. L. Tabernero, N. Verdaguer, M.Coll, I. Fita, G.A. Van Der Marel, J.H. Van Boom, A. Rich and J. Aymami, *Biochemistry 32*, 8403 (1993).
40. K.J. Addess and J. Feigon, *Biochemistry, 33*, 12386-12396 (1994).
41. J.J.P. Stewart, *J. Computer-Aided Molec. Design 4*, 1-105 (1990).
42. W.D. Cornell, P. Cieplak, C.L. Bayly, I.R. Gould, Ir. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldell and P.A. Kollman, *J. Am. Chem. Soc. 117*, 5179-5197 1995).
43. D.A Pearlman, D.A. Case, J.W. Caldell, W.S. Ross, III, T.E. Cheatham, D.M. Ferguson, G.L Seibel, U.C. Singh, P.K Weiner and P.A. Kollman, AMBER *4.1*, University of California, San Francisco (1995).

44. K. Gallo, C. Huang, T.E. Ferrin and R. Landgridge, *Molecular Interactive Display and Simulation* (MLDASplus), University of California, San Francisco (1989).
45. L.I. Kozman, G.G. Gause, V.I. Galkin and Yu.V. Dudnik, *Antibiotics 23,* 602-606 (1978).
46. R.L. Petrusek, G.L. Anderson, T.F. Garner, O.L. Fannin, D.J.Kaplan, S.G Zimmer and L.H. Hurley, *Biochemistry. 20,* 1111-1119. (1981).
47. D.E. Graves, M.P. Stone and T.R. Krugh, *Biochemistry 24,* 7573-7581 (1985)
48. G.V. Gursky and A.S. Zasedatelev, *Sov. Sci. Rev. D. Physico-Chem. Biol. 5,* 53-139 (1984).
49. G.V. Gursky, S.L. Grokhovsky, V.A. Nikolaev, A.N. Surovaya, B.K. Chernov, V.F. Pismensky, N.Yu. Sidorova, A.S. Zasedatelev, A.L Zhuze and R.H. Shafer, *J. Biomol.* Str. Dyn. 12, a081 (1995).
50. G.V. Gursky, V.G. Tumanyan, A.S. Zasedatelev, A.L. Zhuze, S.L Grokhovsky and B.P. Gottikh, *Mol.Biol. Reports* 2, 413-425 (1976).
51. Y. Kim, J.H. Geiger, S. Hahn and P.B. Sigler, *Nature 365,* 512-520 (1993)
52. J.L. Kim, D.B. Nikolov and S.K. Burley, *Nature 365,* 520-527.(1993)
53. G.M. Church, I.L. Sussman and S-.H. Kim, *Proc. Nat. Acad. Sci. U.S.A. 74,* 1458-1462 (1977).

Date Received: June 7, 1996

Communicated by the Editor Valery Ivanov